

The diagram illustrates the flow of information from the Internet to a user's computer. At the top, an oval labeled "INTERNET" is connected by a vertical line to a horizontal bar labeled "FILTERING SOFTWARE". On either side of the filtering software bar are two boxes labeled "PICS". Arrows point from the "INTERNET" oval down to the "PICS" boxes. From the left "PICS" box, an arrow points to a box labeled "HARMFUL INFORMATION". From the right "PICS" box, an arrow points to a box labeled "HARMLESS INFORMATION". A starburst symbol is positioned between the two "PICS" boxes, indicating a point of conflict or decision. An arrow points from the "FILTERING SOFTWARE" bar down to a computer icon at the bottom, representing the final destination of the filtered information.

FIG.2

PRIOR ART

```

<HTML>
<HEAD>
<TITLE>
Example of RSAC1 and SafeSurf Rated Page
</TITLE>

<META http-equiv="PICS-Label" content- '(PICS-1.0
"http://www.rsac.org/ratingav01.html"
1 gen true comment "RASC1 North America Serber" by
"marks@icrosys.com" for "http://www.samplesite.com" on
"1996.09.05T02:15-0800" exp "1997.09.05T02:15-0800"
r (n 2 s 0 v 1 1 2))'>

<META http-equiv="PICS-Label" content- '(PICS-1.1
"http://www.classify.org/safesurf/" 1 gen t for
"http://www.samplesite.com/"
r (SS~~000 2 SS~~100 1))'>

</HEAD>
<BODY>
...

```

FIG.3
PRIOR ART

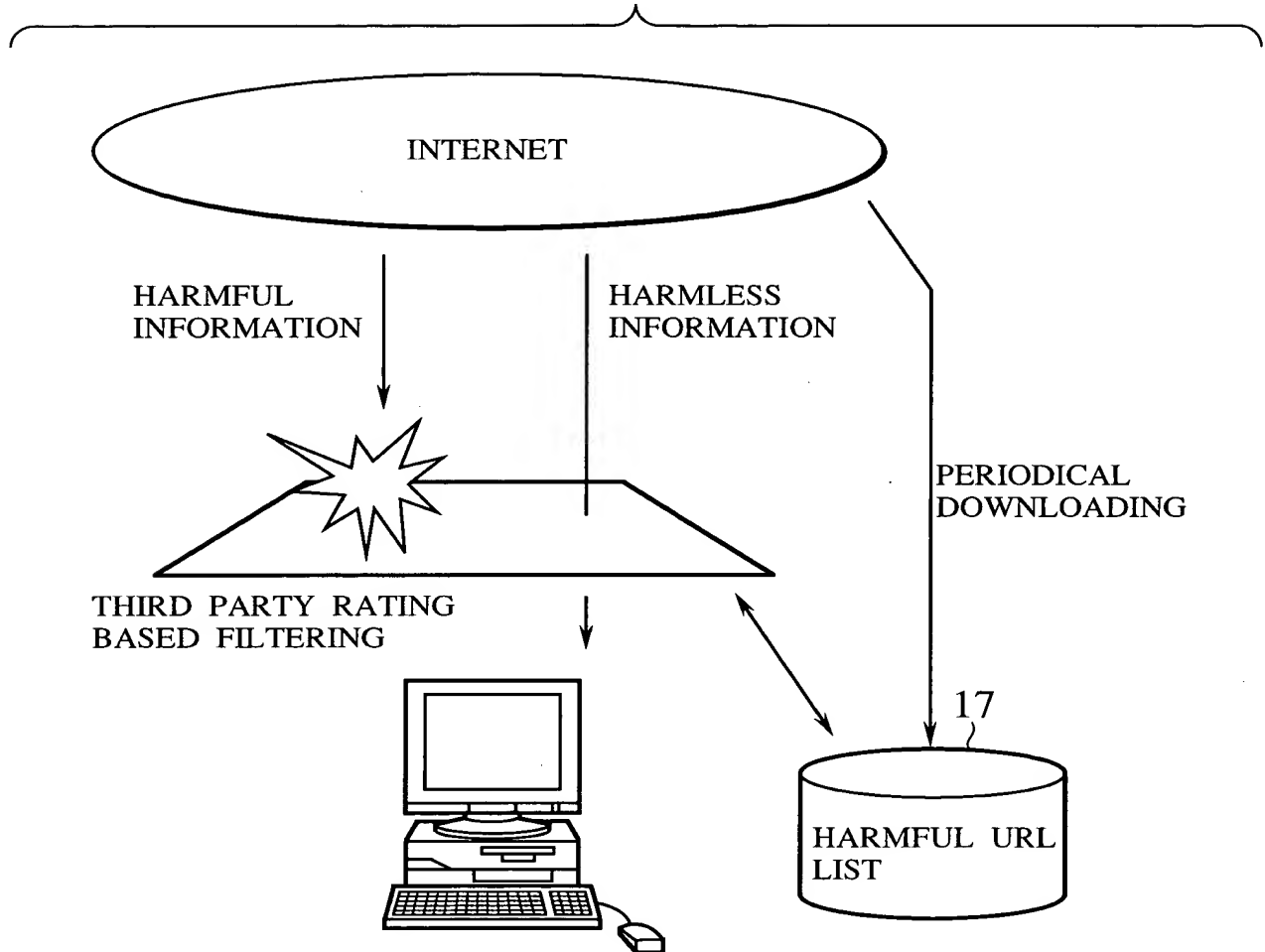


FIG.4
PRIOR ART

PRECISION & RECALL OF VARIOUS KNOWN FILTERING SCHEME

SCHEME	PRECISION	RECALL
SELF-RATING	LOW	LOW
THIRD PARTY RATING	HIGH	LOW
AUTOMATIC FILTERING	LOW	HIGH

FIG.5

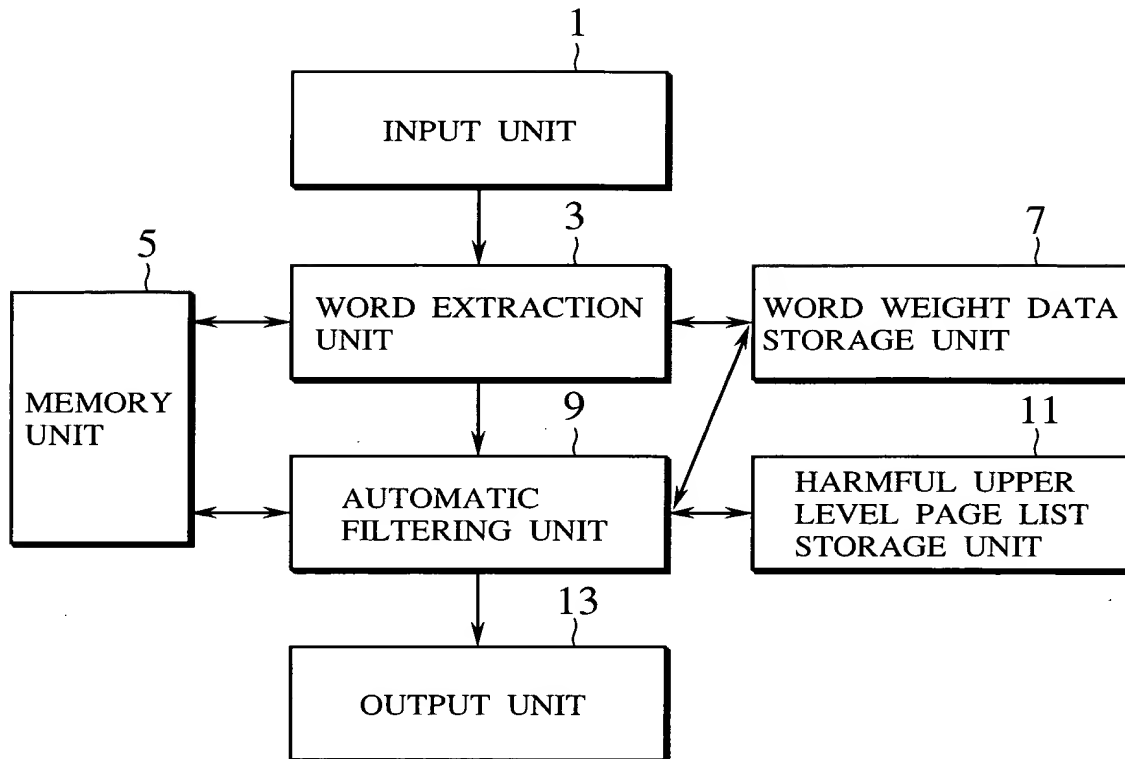


FIG.6

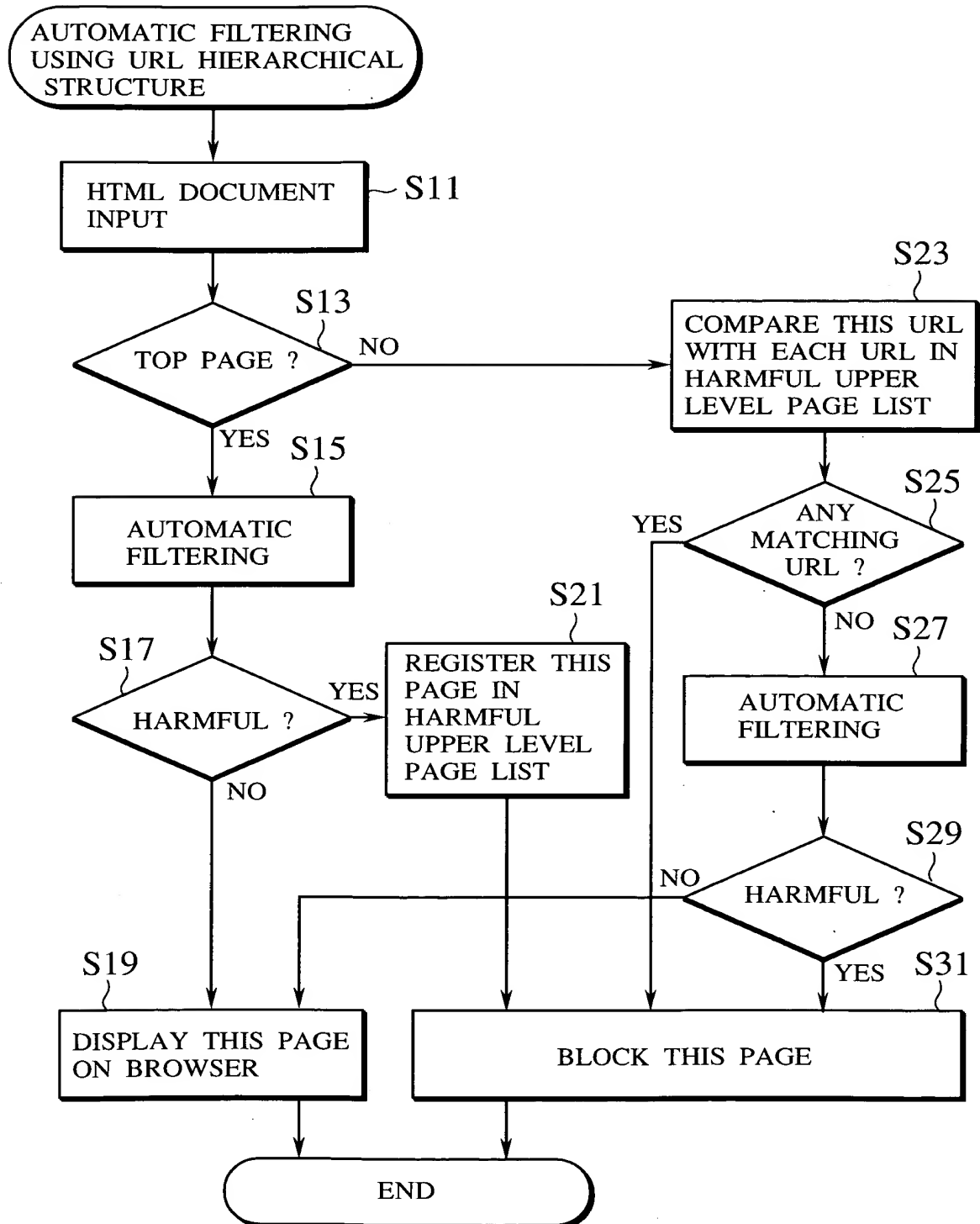


FIG.7

HTML PAGE HARMFULNESS RATING STANDARD

LEVEL	RATING	DESCRIPTION
1	HARMLESS	PAGE WITH NO SEXUAL EXPRESSION
2	PARTLY HARMFUL	PAGE WITH NO DIRECT SEXUAL EXPRESSION BUT CONTENT IMPLYING SEXUAL EXPRESSION
3	HARMFUL	PAGE CONTAINING SEXUAL EXPRESSION

FIG.8

HTML DATA HARMFULNESS RATING RESULT

LEVEL	NUMBER OF DOCUMENTS	RATE
1	6230	22.2%
2	2055	7.3%
3	19749	70.4%

FIG.9

HTML DATA DESCRIPTION LANGUAGE SURVEY RESULT

LANGUAGE	NUMBER OF DOCUMENTS	RATE
ENGLISH	15448	55.1%
JAPANESE	10839	38.7%
OTHERS	1747	6.2%

FIG.10

RATES OF UPPER & LOWER LEVEL PAGES OF HTML DATA IN HARMFUL DATA

TYPE	NUMBER OF DOCUMENTS	RATE
UPPER LEVEL DATA	1495	6.9%
LOWER LEVEL DATA	20309	93.1%

FIG.11

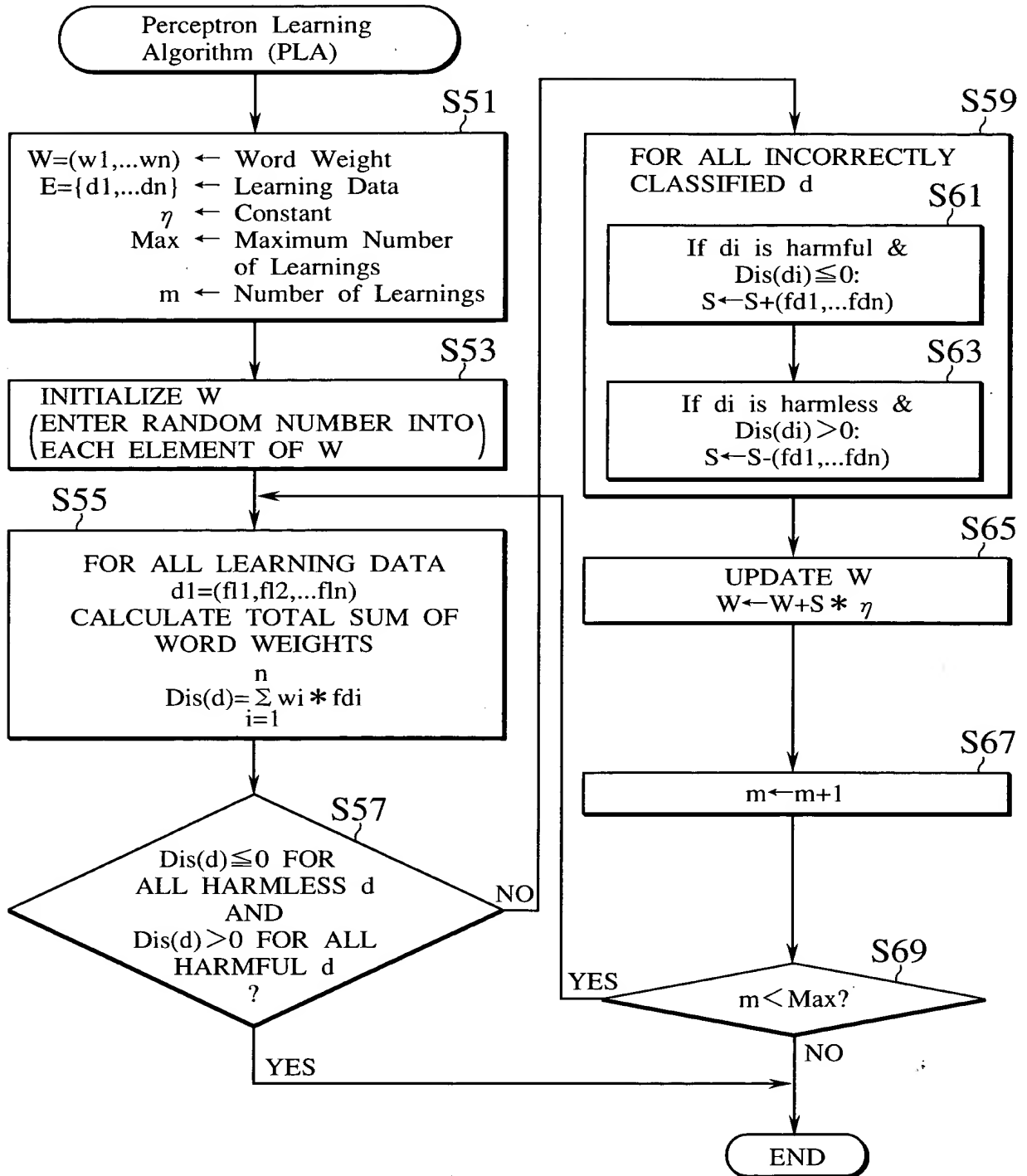


FIG.12

PRECISION & RECALL OF AUTOMATIC
FILTERING ON ALMOST TEXTLESS HTML DATA

min	PRECISION	RECALL
5	88.2%	6.5%
10	92.1%	41.4%
15	91.0%	44.8%
20	91.6%	49.5%

FIG.13

PRECISION & RECALL OF AUTOMATIC FILTERING
ACCOUNTING FOR URL HIERARCHICAL STRUCTURE

min	PRECISION	RECALL
5	91.1%	68.2%
10	91.7%	79.8%
15	90.1%	80.7%
20	90.2%	82.0%

FIG.14

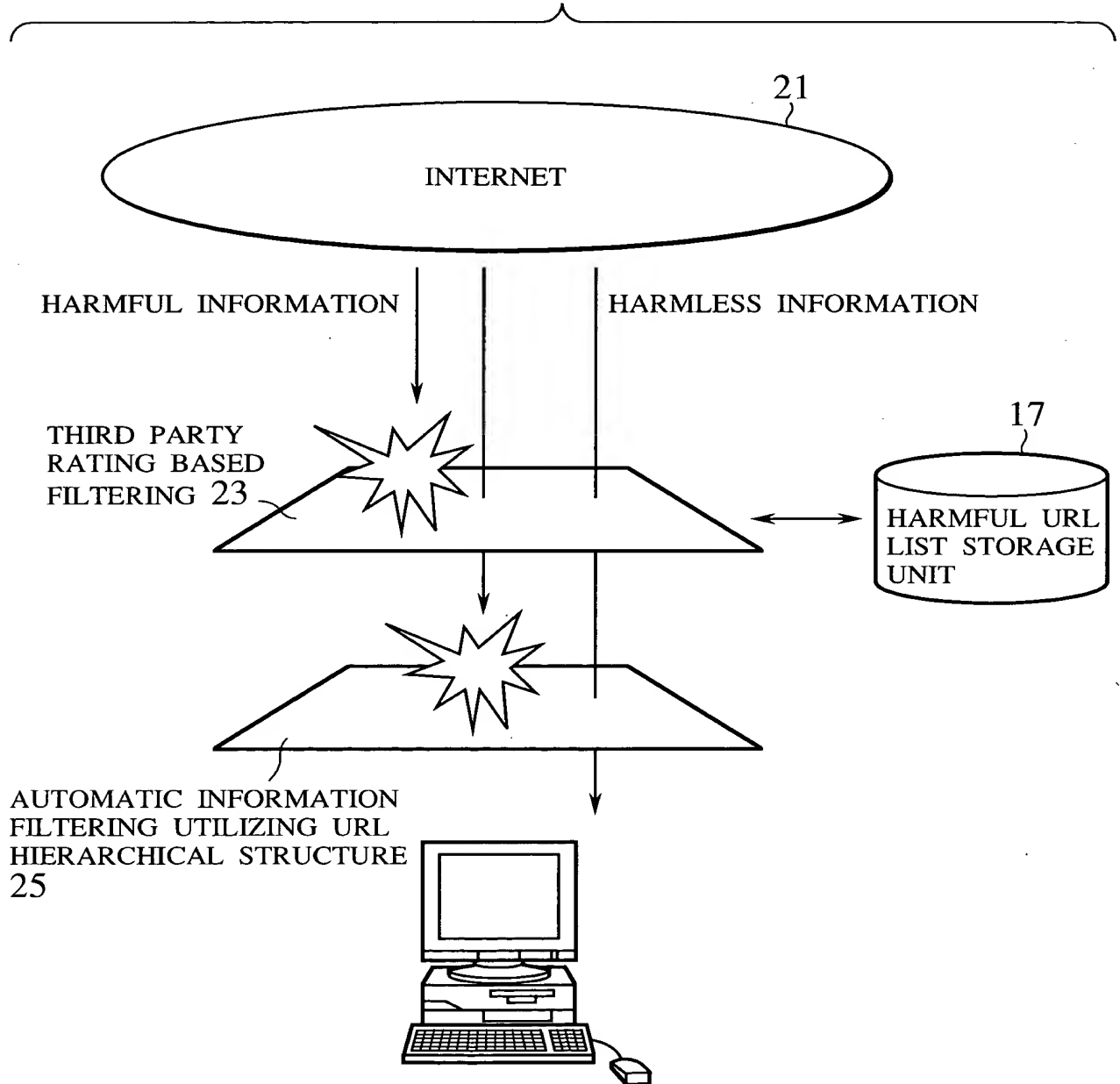


FIG.15

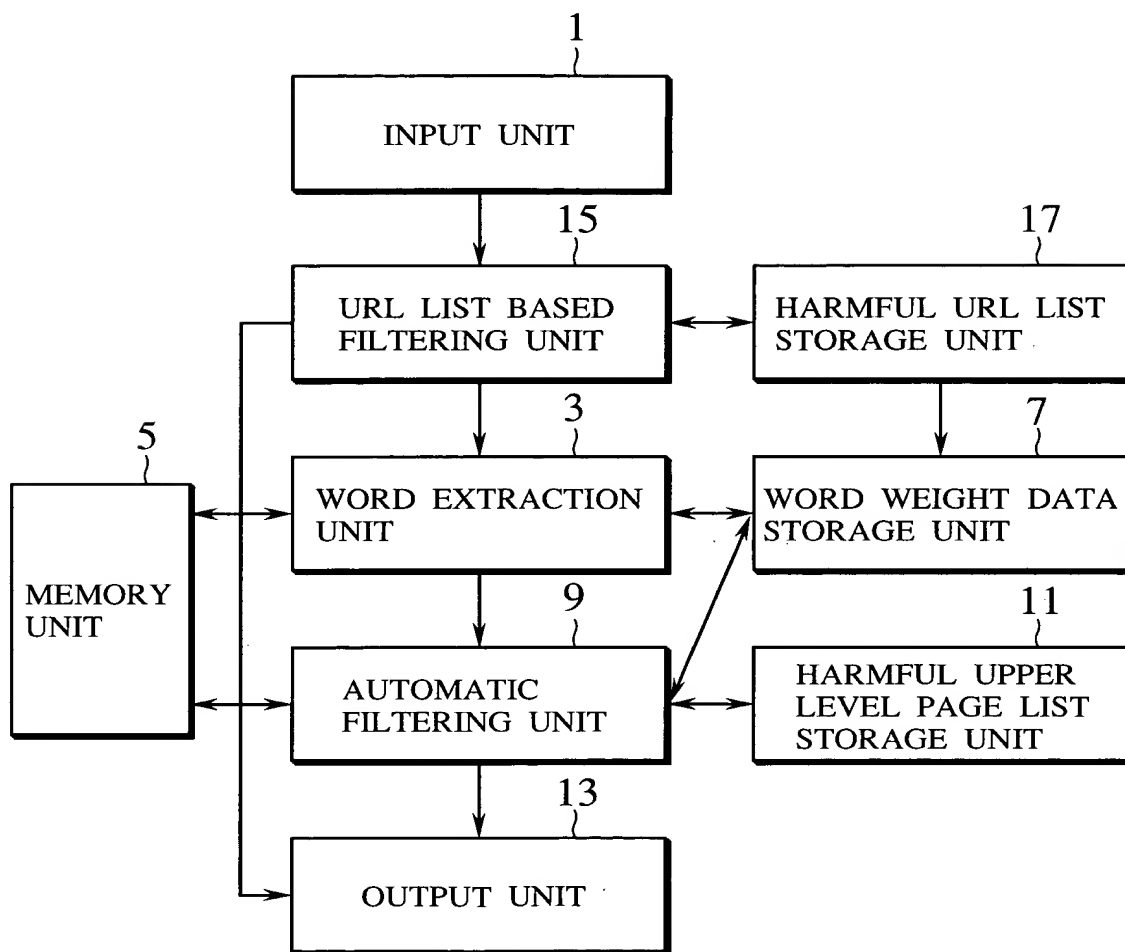


FIG.17

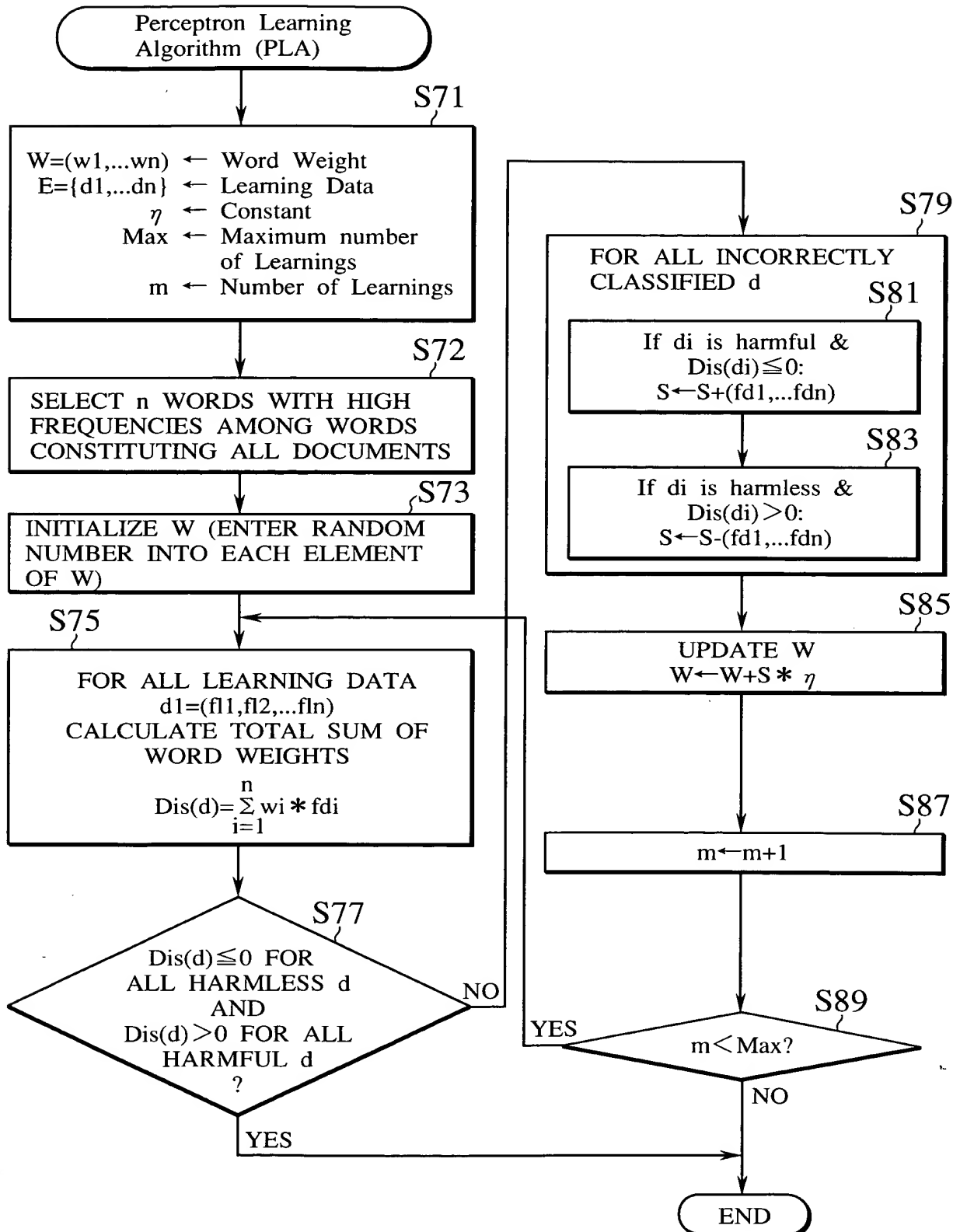


FIG.18

WEIGHTED WORD LIST	
WORD	WORD WEIGHT w
⋮	
IMAGE	10.9
⋮	⋮
SAMPLE	18.7
⋮	⋮
ACCIDENT	-16.6
⋮	⋮
HIGH SCHOOL GIRL	82.2
⋮	⋮
BUS	-101.9
⋮	⋮
HOKKAIDO	-112.5
⋮	⋮
FREE	-6.3
⋮	⋮

FIG.19

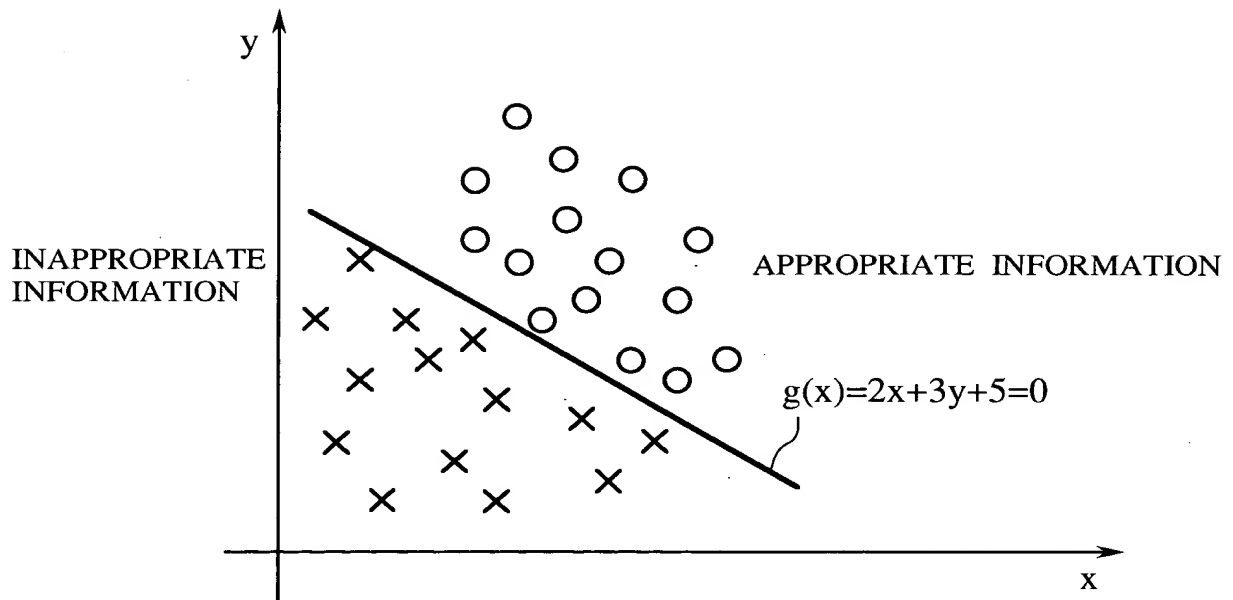
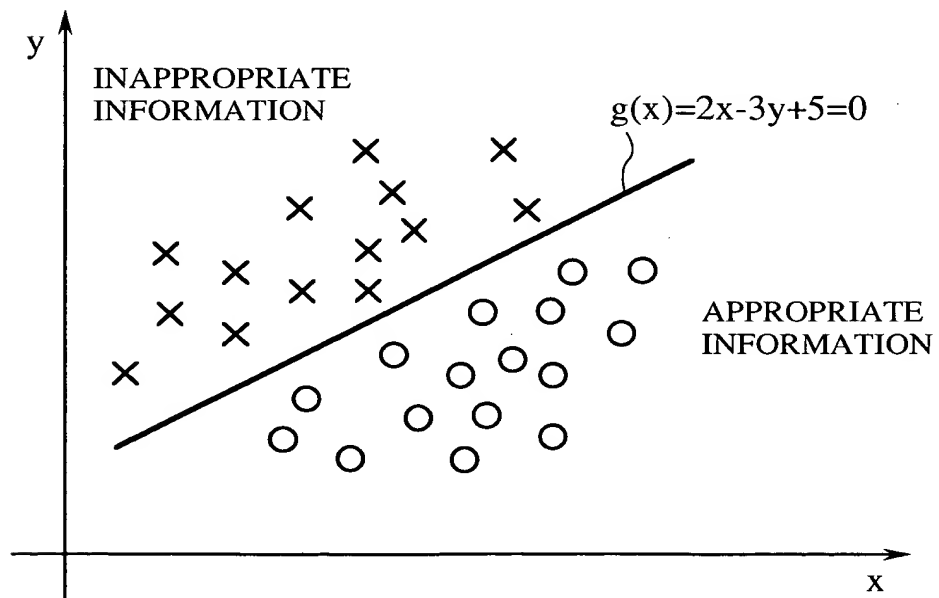


FIG.20



The diagram illustrates a harmful information filtering system architecture. At the top, an oval labeled "INTERNET" (71) is connected to a central processing unit (73) via two vertical lines. The left line is labeled "HARMFUL INFORMATION" and the right line is labeled "HARMLESS INFORMATION". The central processing unit (73) is a trapezoidal shape with a starburst symbol on its left side, indicating a filtering or blocking action. It is connected to a "HARMFUL URL LIST STORAGE UNIT" (77) via a double-headed arrow. Below the central processing unit, another trapezoidal shape represents "AUTOMATIC INFORMATION FILTERING USING AUTOMATIC WORD WEIGHT LEARNING" (75), also featuring a starburst symbol. This unit is connected to a computer icon at the bottom, which consists of a monitor, a base unit, a keyboard, and a mouse.